(2)

EFFECTS OF SELF-ASSESSMENT ON RETENTION OF TRAINING

BY

HEATHER J. S. CABIGON, M.A.

A Thesis submitted to the Graduate School

in partial fulfillment of the requirements

for the Degree

Master of Arts

Major Subject: Psychology

New Mexico State University

Las Cruces, New Mexico

May 1993

May 1993

Effects of Self-Assessment on Retention of Training

Heather J. S. Cabigon

New Mexico State University
Las Cruces, New Mexico

Army Soldier Support Center
Fort Benjamin Harrison, IN

Approved for public release; Distribution Unlimited

See page V

Self-Assessment

Unclassified Unclassified Unclassified Unlimited

"Effects of Self-Assessment on Retention of Training," a thesis prepared by Heather Jeanne Shepley Cabigon in partial fulfillment of the requirements for the degree, Master of Arts, has been approved and accepted by the following:

L L Ames

Lynford L. Ames
Dean of the Graduate School

Darwin P Hunt

Darwin P. Hunt
Chair of the Examining Committee

27 April 1993

Date

Committee in Charge:

Dr. Darwin P. Hunt, Chair

Dr. Dennis B. Beringer

Dr. Melchor Ortiz, Jr.

Dr. Laura A. Thompson

ACKNOWLEDGMENTS

I wish to thank Dr. Darwin P. Hunt for the use of his self-assessment software program and his unwavering dedication to the pursuit of science.

I wish to thank Dr. Melchor Ortiz, Gloria Maese, and the majority of the Experimental Statistics Department for their dedication, patience, and support. Without their assistance this project would have been infinitely more painful.

I wish to thank Mr. Al Moreno for his expert and patient assistance in times of mechanical distress. His ability to debug, rewire, and trouble-shoot complex computer systems was invaluable.

Finally, I wish to thank my friends and loved ones for their support and unshakable faith in my abilities.

VITA

March 4, 1959 - Born at Springfield, Massachusetts

1981 - B.A., University of Massachusetts, Amherst

1981 - present - Officer in the United States Army

1990 - M.A., Webster University, St. Louis

PROFESSIONAL AND HONORARY SOCIETIES

Association of the United States Army

Human Factors Society

ABSTRACT

EFFECTS OF SELF-ASSESSMENT ON

RETENTION OF TRAINING

BY

HEATHER J. S. CABIGON, M.A.

Master of Arts in Experimental Psychology

New Mexico State University

Las Cruces, New Mexico, 1993

Dr. Darwin P. Hunt, Chair

This study examined the effects of self-assessment (SA) responding on the acquisition and retention of rule-based learning. SA responding required the learner to indicate the degree of sureness felt in the correctness of each answer by selecting an SA level of 1-5 ("not sure at all" to "extremely sure") before receiving feedback on the correctness of his response.

The current study found that males scored significantly higher than females on the initial retention trial, lending support to the contention that males perform better on multiple-choice tests.

Analysis of learning and retention trials revealed that learners who self-assessed required significantly fewer trials to meet both the 80% correct learning criterion and the 100% correct relearning criterion than learners that did not self-assess. Learners who did not self-assess, but had knowledge-of-results of the correctness of the response delayed required significantly fewer trials to reach the 100% correct relearning criterion than learners that did not self-assess.

These findings indicate that self-assessment facilitates the acquisition of knowledge and skills, which is beneficially reflected in the correctness of their subsequent performance. The use of self-assessment during learning may result in a higher associative strength between the stimulus and response at the end of the learning phase (reaching the 80% correct criterion). This stronger association may facilitate retrieval and in conjunction with the use of self-assessment during relearning trials, result in fewer trials to reach the relearning criterion (100% correct).

TABLE OF CONTENTS

LIST OF TABLES

LIST OF APPENDIX TABLES

LIST OF FIGURES

LIST OF APPENDIX FIGURES

INTRODUCTION

Hunt (1982) discussed the effects of self-assessment as follows:

> It is widely accepted that human performance is affected by the knowledge that an individual has stored in memory, by the rapidity and accuracy with which such knowledge can be retrieved and processed, and by whether the motor responses required to translate a decision into action can be appropriately selected and executed. The main point of this article is that the processes by which such knowledge and skills are learned are affected by the person's self-assessments of their own performance. (p. 75)

The present study was designed to determine the effect of self-assessment on the learning and retention of the appropriate response.

Self-assessment (as a variation of confidence weighting) is highly documented as a method to improve learning and performance (Echternacht, 1972). The preponderance of research suggests an internal analytical model in which potential responses to the stimulus are evaluated against their consequences and the response that best fits the goal is executed (Hunt, 1984; Kohlers & Palef, 1976; Shaughnessy, 1979).

Kulhavey, Yekovitch and Dyer (1976) propose that learners create a hierarchy of confidence in the correctness of possible choices and then make their response selection based on what they believe to be the most probable right answer. Self-assessment of the correctness of the selected response may

provide an extra source of reinforcement and feedback, which facilitates learning and retention. This study focuses on the use of self-assessment as a reinforcer through its role as feedback in the information processing cycle. Reinforcement is formally accomplished through the rewarding of extra points for accurate self-assessment that the response is correct. Actual values are shown in Appendix F, Appendix Table F-1.

Confidence Weighting and Self-Assessment

Confidence weighting is a method of testing where weights are assigned directly or indirectly to item responses in such a way as to reflect the examinee's belief in the correctness of the alternative or alternatives marked (Echternacht, 1972). The distinction has been made between two different types of confidence testing: confidence weighting and probabilistic testing (Jacobs, 1971). As self-assessment deals primarily with confidence weighting theory, this researcher will define but not discuss the properties of probabilistic testing.

Probabilistic methods require an examinee to assign numbers to each alternative response to a question that reflect the examinee's assessment of the probability of that particular alternative being correct. It is generally required that the probabilities for each item sum to one (Shuford, Albert, & Massengill, 1966).

The vehicle for confidence testing has historically been the multiple-choice test. The multiple-choice test is simple to administer, to take and to score. The simple response format allows more information to be tested in a shorter period of time than do more complicated methods. This test allows objective scoring along a vast range of topics and is used extensively for diagnostic testing and prediction by the American Educational Testing Service. An example of this is the Scholastic Aptitude Test (SAT) which is used in determining college admissions and predicting freshman year success in college.

Confidence testing has a history that dates back to the early 1900s (Henmon, 1911; Hollingsworth, 1913; Trow, 1923). From the 1930s (Hevner, 1932) to the present (Costermans, Lories, & Ansey, 1972) it has been used to increase the reliability of multiple choice testing. Anderson (1982) enumerated several advantages of confidence testing over standard scoring methods. They are as follows:

1. It is well established that increased reliability of testing can be achieved.
2. There is some evidence that examinees pay more attention to multiple choice alternatives.
3. More diagnostic information becomes available.

4. Pre- and post-examination tension is reduced, resulting in happier test takers.

According to Anderson (1982) confidence weighting methods usually require examinees to reveal their degree of confidence in the correctness of each conventional response by choosing among two or more confidence scale points, each identified by a general verbal description. Hunt's (1982) scale of "sure" and "not sure" is representative.

Many different scoring functions have been used in confidence weighting. Across the board they have rewarded proportionate confidence in the correct answer and some have punished confidence in the wrong answer (Hunt, 1982).

Proponents of confidence testing claim that "two aspects of knowledge are what one believes to be true and how confident one is in that belief" (Fischoff, Slovic, & Lichenstein, 1977). Knowledge is a continuous stream and can not be partitioned into what is known and what is unknown. There is partial knowledge, although conventional scoring functions work best when partial knowledge does not exist (Abu-Sayef, 1979). Rippey and Voytovich (1983) state that "in the face of incomplete knowledge we are often inept and uncomfortable even though partial knowledge can be better than no knowledge at all." Confidence testing acknowledges the value of partial knowledge and gives credit for it.

Hunt (1984) developed a theory of the human self-assessment process based on the history of the effects of confidence weighting, his personal history as an educator, and his own experimental findings. Hunt contends that people make self-assessments of the state of their own subsystems and these assessments affect their decisions, and the selection and execution of responses. His experimental findings show that self-assessment affects the processes by which knowledge and skills are learned and the correctness of their subsequent performance. Hunt (1978; 1982) found that not only was the acquisition of a paired-associates learning task facilitated, but also the response time was shortened through self-assessment. In the same learning task, correct answers about which a respondent was "sure" were executed 1.5 seconds faster than answers about which the respondent was "sure" but which were wrong.

Both Hunt's (1984) Eight-Factor Model of Human Performance and Learning and the revised 1989 version attempt to describe the internal psychological learning processes as intervening variables (Figure 1). This is accomplished by defining the functions of each of the components and their interactive channels and operations. The focus is on the response or output.

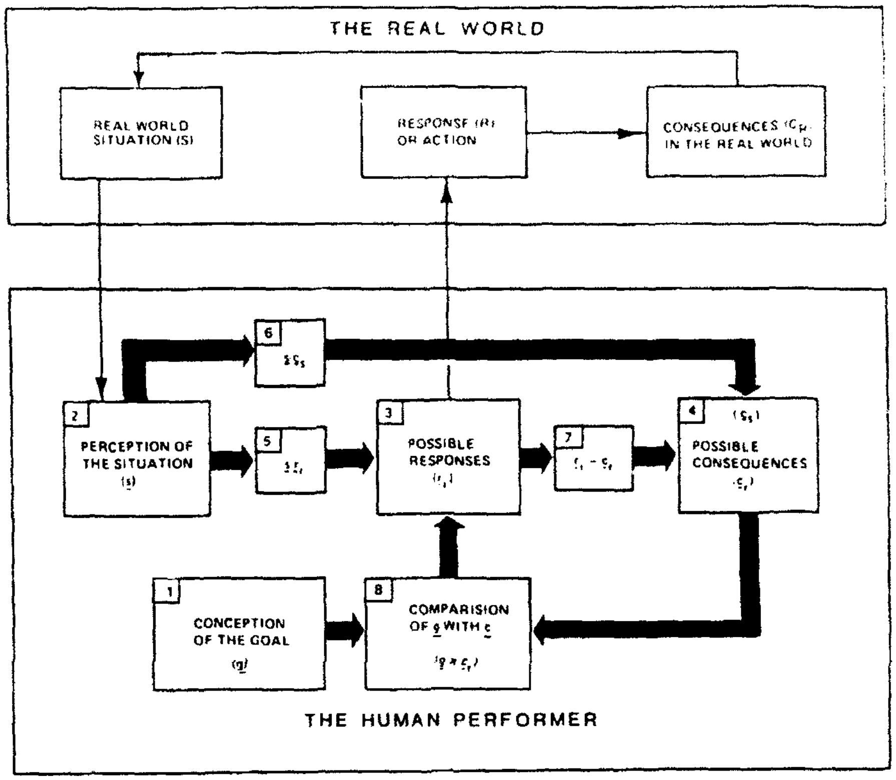


Figure 1. An eight-factor model of human performance and learning (Hunt & Sams, 1989, page 45).

The researcher will attempt to explain these components by walking through the model using the scenario of an experimental subject taking a multiple-choice test with self-assessment feedback and knowledge-of-results feedback on the correctness of the response.

The real world situation is that the learner must get this question correct if he is to reach the criterion of 100% correct on the test, receive his experimental credit and be dismissed. The learner accurately perceives this situation and formulates the immediate goal of executing the correct response.

The learner processes the possible responses with their subsequent consequences (correct or incorrect) and compares each to the goal of executing the correct response. The learner then selects the response that he feels best meets the goal. The strength of the learner's belief that he executed the best response (correct) to meet the goal of answering the question correctly is the basis for his self-assessment of the degree of sureness that the chosen response is correct. The feedback the learner receives on the accuracy of his self-assessment and the correctness of the chosen response, provide the information necessary to modify his internal model.

Hunt's eight-factor model can be used as the conceptual outline for the process of rule-based learning through performance and reinforced through feedback. The individual is exposed to the rule-based information and places it in memory in accordance with individual capabilities. The details of the learner's internal model are formulated based on this

perception. Input in the form of a particular stimulus is perceived and the model is activated. The chosen response is executed with the expectation that it best meets the goal. Knowledge-of-results provides the information necessary to modify the internal model and meet the goal with increased accuracy.

Evaluation of learning is based on the correctness of responding and retention when tested. To accomplish the required task correctly, the learner must have accurately developed, activated, and tailored his internal model. Self-assessment is part of the process by which this is accomplished.

Hunt's formalization of internal memory trace procedures and analysis in his Eight Factor Model of Human Performance and Learning is supported by other researchers findings. Shaughnessy (1979) states that current theories of memory place heavy emphasis on the learner's sensitivity to internal feedback concerning the quality of ongoing encoding activities. Memory monitoring processes serve as the foundation for viewing the learner as a decision maker who selectively draws upon a repertoire of mental devices depending upon the perceived condition and task demands. This aligns with Hunt's view of the learner actively using feedback in the form of self-assessment and response accuracy to modify his internal model and increase the probability of execution of the correct response to best meet the perceived goal. Kohlers and Palef (1976) research is consistent with Hunt's

theory, viewing learning and retention as an internal, procedure based, analytical model.

Kohlers and Palef (1976) propose that recognition can occur as the reinstitution of the set of internal analytical operations that were first activated as the reaction to a stimulus. Those operations have a structure, and particular order, so that some are carried out contingently on the execution of others. Kohlers and Palef close by stating emphatically that recognition is not a matter of matching a stimulus to a stored representation or memory trace but is due to a set of procedures or analytical operations that are carried out to some variable criterial level.

By evaluating the effects of active self-assessment during both the learning and relearning trials on the percent of correct responses, the researcher hoped to determine whether (a) self-assessment improves retention through utilization as a reinforcer in the feedback loop, and (b) the percentage of correct responses about which the learner is sure are retained more frequently than are correct responses about which the learner is unsure.

Self-Assessment as a Reinforcer

Historically, learning theory has been divided into two camps, the cognitive and the stimulus-response (or associationist) theorists. The associationist approach is of primary interest to this research.

The associationist theorists such as Thorndike (1898), and Hull (1943; 1952) have proposed that learning cannot take place unless behavior is reinforced during the learning process. Contiguity alone between stimulus and response is not enough.

Thorndike explained the instrumental learning he witnessed in his animal experiments through the Law of Effect. Essentially, this stated that the subject has many potential responses and some of these responses occur during the learning situation. In time, one of these responses leads to a reward and is strengthened. This increases the probability of it occurring again and being strengthened again. With enough rewarded repetitions, the response will occur at a reliable rate and is chosen from among all potential responses.

Hull (1943) follows the tradition of Thorndike by viewing learning as the establishment of associations between stimuli and responses and regarding the effects of behavior as the crucial event in forming these associations.

Hull (1952) analyzed behavior in terms of the elemental stimulus-response units that characterized conditioning experiments. He stressed the external processes in learning as the modification of behavior through the development of associations between stimuli and responses.

Organisms learn to behave certain ways in certain situations. These broad, elegant theories laid the groundwork for modern learning theories.

It is under this definition of reinforcement's role in learning that self-assessment can be viewed as a reinforcer, particularly as a form of knowledge-of-results. This knowledge-of-results or feedback has three major roles in learning. It may act as a reinforcer as explained above, it may provide information to the learner, or it may provide motivation.

Self-assessment acts as a reinforcer by providing knowledge-of-results in the feedback loop. This information on the examinee's confidence in the correctness of the chosen response assists in the initial learning process.

Self-Assessment as Feedback

Performance improves when knowledge-of-results is presented immediately after a response (Anderson, Kulhavy, & Andre, 1971; Meyer, 1960). This improvement is viewed under both an associative and motivational methodology (Hogarth, Gibbs, McKenzie, & Marquis, 1991). The associative view is that knowledge-of-results acts as a directive or guiding influence so the organism learns which stimulus-response pairings are correct.

The motivational view sees knowledge-of-results as an energizing influence so the organism strives to make more of the responses already available to it.

Feedback serves as a source of information that provides more than simple reinforcement. Carroll (1976) states that feedback not only confirms and reinforces correct responses, but disconfirms and provides information to correct wrong responses. The extent to which informative feedback improves performance relies on effective processing by the learner. Hunt (1982) concluded that the confirmation of correct responses was of fundamental importance in learning. He found that the hit rate, but not the false alarm rate, was higher in those treatments which enhanced learning.

Phye and Bender (1989) agree that feedback is a closed-loop system where the learner desires to minimize the deviation between the actual and desired value, the desired value in this case being a perfect score on a exam. They cite Fisher and Mandl (1985) who argue that this self-regulating system possesses the competence to regulate itself. The learner is not only competent to self-regulate, but the self-regulation proceeds automatically without conscious thought and does not strain capacity to the point of interfering with the original task. This is remarkably similar to Hunt's (1984)

Eight Factor Model of Learning. The learning effects of self-assessment as a form of feedback are demonstrated in the following research.

Kulhavey et al. (1976) ran a study in which 60 college students read and were tested on a program about the human eye. Subjects either did or did not receive feedback regarding the correctness of the answer after each response. All subjects self-assessed on the degree to which they felt each answer was correct. Each subject was post-tested immediately and tested again after one week. The effects of feedback were strongly influenced by the learner's perception of the accuracy of his response.

Kulhavey et al. (1976) concluded that feedback provided the biggest benefit when it followed a high confidence response. They found that when the response was correct, seeing feedback increased the probability of repeating the right answer, particularly on the immediate post-test. When confidence was high and the response was incorrect, feedback provided a strong corrective device. This effect maintained itself over a retention interval. Results of time-on-feedback analysis support Kulhavey et al. in their contention that correction of high-confidence errors occurs because subjects spend more time reprocessing the information. This dramatic rise in attending behavior increases the chances that the error will be corrected on later tests.

Kulhavey et al. neglect to mention that the increase in attending behavior could be caused by the "surprise" effect of high confidence in an incorrect response. According to Johnston (1979), the condition under which learning occurs is cortical arousal or desynchrony. This is reflected in large P3 waveforms of the evoked potential. Events that have high utility (reinforcing) and high information (novel) have been shown to produce large P3 waveforms by Sutton, Braren, Zubin, and John (1965). This phenomena is also cited by Crouse (1980) in his unpublished thesis, but his findings did not support the surprise effect. As applied to Hunt's (1984) Eight Factor Model of Human Performance and Learning, the learner is surprised at the inaccuracy of his internal model and spends more time on the feedback to correct the model and bring it in line with the new information.

Hypothesis

Self-assessment has been shown to facilitate learning (Hunt, 1982; Sams, 1986). Confidence weighting has a long history as a method to improve learning and performance (Echternacht, 1972). Performance improves when knowledge-of-results is presented immediately after a response (Anderson et al., 1971; Meyer, 1960). Knowledge-of-results or feedback functions as a source of information that provides more than simple

reinforcement. Carroll (1976) states that feedback not only confirms and reinforces correct responses, but disconfirms and provides information to correct wrong responses. Therefore, it is suggested that self-assessment acts as a form of knowledge-of-results and meets the definition of a positive reinforcer of learning. Positive reinforcers strengthen the stimulus-response association and increase the probability the response will recur (Skinner, 1938). Those items with the strongest stimulus-response association as indicated by the self-assessment level, should produce the greatest retention (Kulhavey et al., 1976).

It is hypothesized that those correct responses about which a person is "extremely sure" at the time of acquisition will be retained at a higher percentage than those correct responses about which the person is "unsure."

Pilot Study

To evaluate and develop the experimental procedures, a pilot study was conducted. Subjects were required to learn and apply the rules for identifying and extinguishing four classes of fire to a specified level of performance. The experimental group overtly self-assessed the confidence they felt in the correctness of their response. This self-assessment was immediately followed by feedback on the correctness of the response. The control group was not

required to self-assess and their response was immediately followed by feedback on the correctness of the response. Learning was measured by the number of correct answers on a 30 question multiple choice task. After achieving an 80% correct criterion during the learning session, the subject was required to return one week later on the same day of the week and at the same time of day to take an identical multiple choice task to evaluate retention. See Appendix A for more detailed information regarding the pilot study.

An analysis of variance procedure using total number correct as the dependent measure revealed a significant first order interaction [$F(2, 19) = 4.1$, $p < .05$] between grade point average (GPA) and ethnicity. Protected Least Significant Difference analysis indicated an effect of GPA level on mean retention scores for White Non-Hispanics ($p \leq .05$) but little or no effect for Hispanic/Native American mean retention scores ($p > .05$).

Although there were not enough observations for statistical analysis, there appeared to be a relation between how sure examinees were of the correctness of their answers and the number correct on the retention trial. This effect is the focus of the current research. As demonstrated by Kulhavy et al., (1976), feedback reaps its biggest benefit when it follows a high-confidence response. When the response is correct, receiving feedback

increases the probability of repeating the right answer. When confidence is high and the response is incorrect, feedback acts as a strong corrective device.

Encouraged by the potential for self-assessment to effect acquisition and retention of training, the following changes were made to modify the design and procedures for the current study.

The basic experimental design remained the same as the pilot study. However, an additional treatment group was included to control for confounding of self-assessment data due to the delay in receiving knowledge-of-results caused by the subject taking time to self-assess. This delay-knowledge-of-results group did not self-assess, but knowledge-of-results on the correctness of the response was delayed to approximate the time required to self-assess. This time period was determined by earlier experimentation. To control for any variance due to the effects of ethnicity and gender each group was balanced with equal numbers of Hispanics, White Non-Hispanics, males and females.

Additional data were collected on the relearning session by requiring each subject to continue to complete trials until a criterion of 100% correct was met.

The time allocated for learning and relearning sessions was increased from one hour used in the pilot study to two hours for the current study. This was divided into one hour for the learning session and one hour for the relearning session. This change was made because during the pilot study, subjects generally needed more time for learning trials and the added requirement of a 100% correct criterion for the relearning session would require more time.

METHOD

Subjects

One hundred and twenty undergraduate college students served as subjects to partially satisfy requirements in an introductory psychology course. The data of an additional 32 subjects were unusable as 14 subjects failed to return for the retention portion of the experiment, or 12 were unable to learn the task to the 80% correct criterion, or 6 were unable to relearn the task to the 100% correct criterion. Subjects were randomly assigned to three experimental groups. These groups were the self-assessment (SA) group, the no-self-assessment (NOSA) group, and the no-self-assessment-delayed-knowledge-of-results (DKR) group. The SA group overtly self-assessed the confidence they felt in the correctness of their response. This self-assessment was immediately followed by feedback on the correctness of the response. The NOSA group was not required to self-assess and their response was immediately followed by feedback on the correctness of their response. The DKR group was also not required to self-assess but the feedback on the correctness of their response was delayed to approximate the time required for the experimental group to self-assess.

The primary task of all subjects was to learn the four classes of fire, the rules to determine which type of fire fell under which class, and the type of fire extinguisher required. Learning, retention and relearning were measured by the number correct on a 30 question multiple-choice task. One iteration of this 30 question task was considered one trial. After achieving an 80% correct criterion during the learning session, the subject was required to return one week later on the same day of the week and at the same time to take an identical multiple choice task to evaluate retention and relearning. The retention trial was the first trial of the relearning session. The relearning session was repeated until the subject was correct on all answers. After achieving this criterion, subjects were debriefed and dismissed.

Each group was balanced with equal numbers of Hispanic males (10), Hispanic females (10), White Non-Hispanic males (10), and White Non-Hispanic females (10) for a total of 40 subjects in each group. No other ethnic groups were tested. Ethnic affinity and grade point average (GPA) were self-reported. Testing continued until balanced groups of 40 subjects each were achieved (Table 1).

Table 1

Experimental Design

	MALE HISPANIC	FEMALE HISPANIC	MALE WHITE NON-HISPANIC	FEMALE WHITE NON-HISPANIC	TOTAL N
SA	10	10	10	10	40
NOSA	10	10	10	10	40
DKR	10	10	10	10	40

NOTE: SA = SELF-ASSESSMENT NOSA = NO SELF-ASSESSMENT
DKR = DELAY KNOWLEDGE OF RESULTS

Apparatus

A personal computer was used to present task information, provide instructions, present the learning and testing trials and collect data on subject responses. All relevant aspects of the workstation were in accordance with ergonomic principles determined by The Human Factors Design Handbook (Woodson & Tillman, 1992) and are discussed in Appendix B.

Procedure

Each subject in the three groups participated individually at a separate workstation using a standardized experimentation schedule. Subjects were tested in groups of ten. Sampling was regulated to achieve groups balanced for ethnicity and gender, and counter-balanced for time-of-day and day-of-week (see Appendix C for a representative schedule).

For all groups, subjects were instructed that the task was to learn the four classes of fire, and how to recognize and extinguish each class. They were told to study the information when displayed on the computer screen because a 30 question multiple-choice test would follow and the criterion for learning the material was a score of 80% correct. See Appendix D for detailed instructions.

During both learning and relearning trials, all subjects selected a response (A, B, C, or D) by pressing the corresponding pushbutton on the computer keyboard. The control group received immediate feedback on the correctness of the selected response. The experimental group was first required to express the confidence they felt in the correctness of the selected response by selecting a self-assessment level of 1-5 ("not sure at all" to "extremely sure"). After selecting this level, each subject received feedback on the correctness of the selected response. The delayed-knowledge-of-

results group received feedback on the correctness of the selected response after an interval that approximated the delay caused by self-assessing. This interval was determined by preliminary experimentation to be approximately one second.

Upon achieving a learning score of 80% correct, the subject was instructed to return for the retention and relearning trials the following week on the same day of the week and at the same time. The retention test score was drawn from the first trial of the relearning session. Relearning trials were considered all trials from the start of the relearning session (including the retention trial) through completion of the 100% correct criterion. When subjects successfully achieved the 100% correct criterion they were debriefed and released. A sample of the multiple-choice task is provided in Appendix E. The standard self-assessment format that followed every test question for the self-asscssment group is found in Appendix D in the test instructions.

RESULTS

The percentage of correct responses was calculated for the first retention trial for each individual in the (SA) self-assessment group, the (NOSA) no-self-assessment group, and the (DKR) delay of knowledge-of-results with no self-assessment group. These data were then analyzed through ANCOVA with the following independent variables: Group, Ethnicity, and Gender. The covariate was Grade Point Average. This analysis was required to equalize grade point averages as it was believed that performance on the task might be related to grade point average. The dependent variable was the percent correct on the first retention trial.

The ANCOVA revealed a significant main effect of gender, $\underline{F}$ (1, 119) = 10.26, $\underline{p} < .05$. Males scored significantly higher ($\underline{M}$ = 88.3) than females ($\underline{M}$ = 82.9) on the retention trial. No significant difference was found in performance between treatment groups or ethnic groups. More specific information on the ANCOVA can be found in Appendix G.

The number of learning trials required to reach the 80% correct criterion was compared for the three treatment groups using the chi-square procedure. The number of learning trials was significantly different χ^2 (8, $\underline{N}$ = 120) = 16.3, $\underline{p} < .05$ among treatment groups (Figure 2). The self-assessment group

had more subjects meet the criterion on the first trial than the other two groups. The no-self-assessment group had more subjects achieve the criterion with 4 or more trials than the other two groups. More specific information on the chi-square analysis can be found in Appendix G.

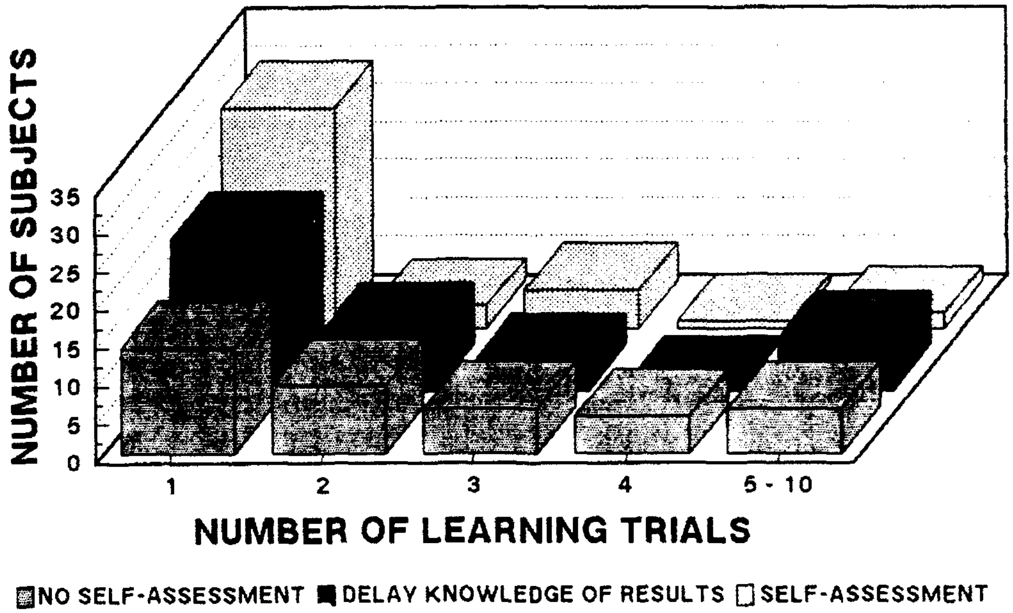
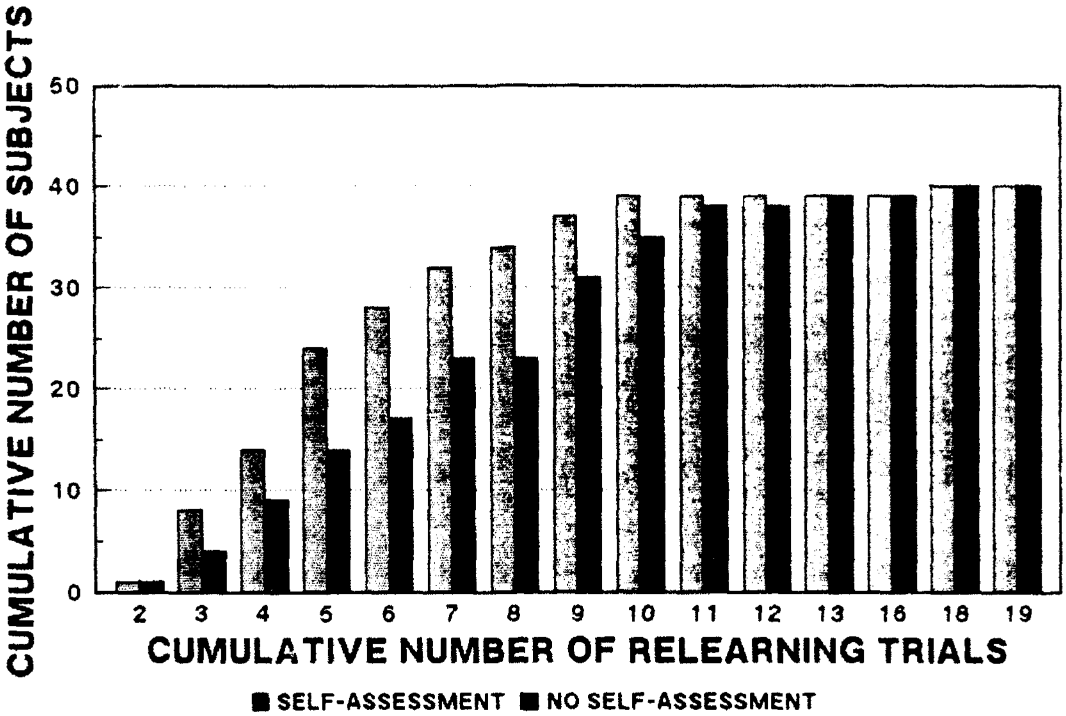


Figure 2. Number of subjects to reach 80% correct criterion by number of trials for each treatment group.

Distribution of the data for the relearning session indicated that the Kolmogorov-Smirnov procedure was the most appropriate (Figure 3). The number of relearning trials required to reach the 100% correct criterion was significantly different ($.025 < p < .05$) among treatment groups. The self-assessment group had more subjects meet the criterion in fewer trials than the no-self-assessment group. The delay-knowledge-of-results with no self-assessment group had more subjects meet the criterion in fewer trials than the no-self-assessment group. There were no significant differences between the self-assessment group and the delay-knowledge-of-results with no self-assessment group. More specific information on the Kolmogorov-Smirnov analysis can be found in Appendix G.

The relationship between (a) the percentage of correct responses on the final learning trial that were also retained correctly on the retention trial and (b) the self-assessment level assigned to those correct responses during the final learning trial was analyzed. The percent of correct responses is shown as a function of level of confidence as indicated by the subject's overt self-assessment during the final learning trial (Table 2 and Figure 4).



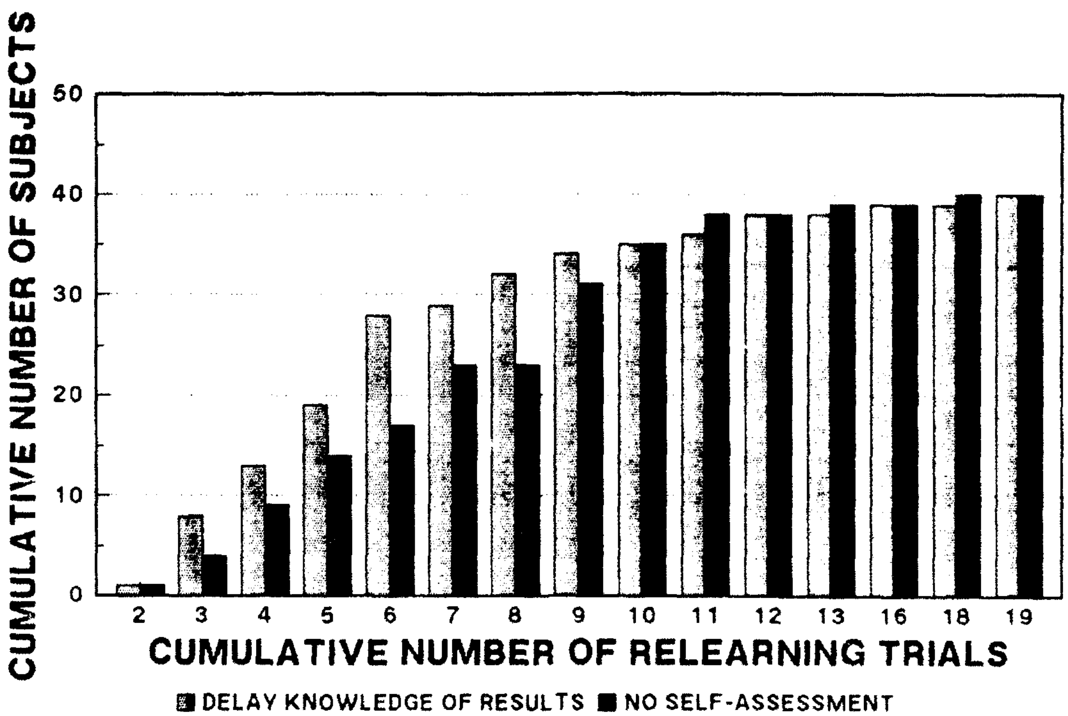


Figure 3. Differences between treatments by cumulative number of relearning trials to reach 100% correct criterion.

The percentage of correct responses on the final learning trial that were also retained correctly on the retention trial was not significantly different among self-assessment levels ($p > .05$). Specific information on the chi-square analysis can be found in Appendix G.

Table 2

Number and Percent of Final Learning Trial Correct Responses Retained Correctly by Self-Assessment Level

SA LEVEL FINAL LEARNING TRIAL	NUMBER OF CORRECT RESPONSES AT EACH SA LEVEL FINAL LEARNING TRIAL	NUMBER OF CORRECT RESPONSES AT EACH SA LEVEL RETENTION TRIAL	PERCENT RETAINED CORRECTLY
NOT SURE AT ALL	4	1	25.0
VERY UNSURE	8	6	75.0
SOMEWHAT UNSURE	53	40	75.5
VERY SURE	97	85	87.6
EXTREMELY SURE	915	835	91.3

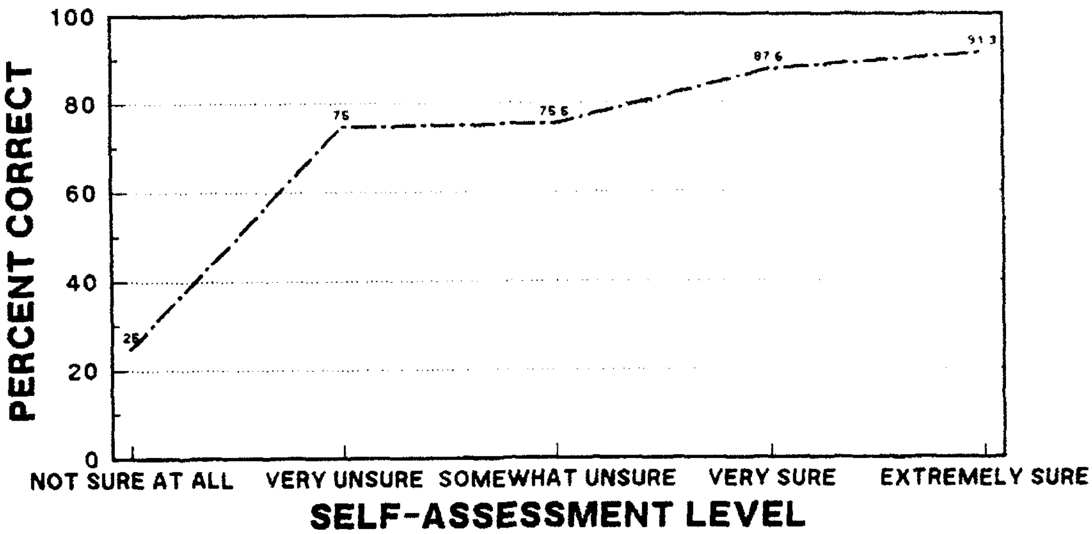


Figure 4. Percent of responses which were correct on final learning trial retained correctly by self-assessment level.

DISCUSSION

These findings indicate that self-assessment facilitates the acquisition of knowledge. The hypothesis that the level of confidence expressed in the correctness of the answer during learning is related to the percentage of answers that are retained correctly is not supported by the data analysis. Further study is required with a large enough subject pool to insure an adequate number of responses in the lower confidence levels to allow for more conclusive analysis.

The results of the ANCOVA seem to support the contention that females do not perform as well as males on multiple choice tests (Hyde, 1981). This finding must be considered in light of the subject matter of the test.

Alice Eagly (1978) makes a strong argument that gender bias is an effect of the experimental subject matter. In this meta-analysis of persuasion and conformity research, Eagly finds that "people tend to conform on matters in which their own sex is thought to be relatively uninterested and inexpert" (p. 97). Eagly feels females' performances suffer because they lack interest and expertise in the experimental subject matter.

Although the same information was presented to each treatment group, males may well have had a higher interest and level of expertise in recognizing and extinguishing the four classes of fire. The females' information-deficit and disinterest in this topic could have resulted in lower scores. Further research contrasting traditionally male topics (politics or economics) with traditionally female topics (nutrition) may better explain this effect.

Analysis of the learning session trials supports Hunt's (1978, 1982) findings that self-assessment facilitates acquisition in a paired-associates learning task. Chi-square analysis found that subjects that self-assessed required significantly fewer trials to reach the 80% correct learning criterion. This is consistent with Hunt's view that the learner actively uses feedback in the form of self-assessment accuracy and response correctness to modify his internal model, thereby increasing the probability of execution of the correct response to best meet the perceived goal.

Analysis of relearning session data found that more subjects that self-assessed reached the 100% correct relearning criterion in significantly fewer trials than subjects that did not self-assess. The group that did not self-assess but had delayed knowledge-of-results also reached the criterion in significantly fewer trials than the no-self-assessment group. However, there

was no significant difference between the self-assessment group and the delay-knowledge-of-results with no self-assessment group. This researcher does not know how to interpret this finding and speculates that the increased length of the pre-knowledge-of-results period (equal to the self-assessment group) allowed the learner to engage his mental model and compare potential responses against their consequences, choosing the response that he felt best met the goal. In effect, the learner was self-assessing and receiving the stimulus-response strengthening benefit of that additional information. This research cannot speculate why this effect is present only during relearning and not during acquisition.

These findings support Carroll's (1976) view that feedback (self-assessment in this study) confirms and reinforces correct responses. This also conforms to Phye and Bender's (1989) characterization of feedback as a closed-loop system, where the learner desires to minimize the deviation between the actual and desired value. Knowledge-of-results (feedback) in the form of response correctness and self-assessment accuracy provides the information necessary to modify the internal model and meet the goal with increased accuracy. This increased accuracy is demonstrated by the self-assessment group requiring fewer trials to reach learning criterion than the groups that did not self-assess.

These results have significant implications for training, particularly when the information being trained must be recalled. Rippey and Voytovich (1983) argue that the increased use of confidence-weighting (self-assessment) during learning can improve performance, especially in programs that require decision making under conditions of incomplete information or uncertainty. Self-assessment gives credit for partial knowledge and helps identify the strength of the stimulus-response association for both correct and incorrect information. It is in this area that self-assessment can have a great impact on training. Self-assessment identifies high confidence correct information and high confidence misinformation. If the criteria for achieving a specified level of training included the self-assessment level assigned to the response rather than correctness alone, it would assist both learner and instructor in identifying the strength of stimulus-response, response-consequence and stimulus-consequence associations (Figure 1) and the existence of possible misinformation. This allows for modification of training programs before incorrect responses are executed, conserving valuable resources such as time and money.

Although chi-square analysis of the percentage of correct responses on the final learning trial that were retained correctly on the retention trial did not show any significant difference among self-assessment levels, there were

indications of differences. The learner retained a progressively higher percentage of correct responses as confidence in the correctness of the response increased. This was similar to the observational results found in the pilot study (Appendix A, Figure A-2), suggesting that this relationship is deserving of further study. Further experimentation with a large enough subject pool (or a lower learning criterion) to provide an adequate number of responses in the lower self-assessment levels (1-3) is required for more conclusive analysis.

Further experimentation is required to examine the parameters of the relationship between level of self-assessment and retention to better facilitate the learning process and maximize training efficiency.

REFERENCES

Abu-Sayf, F.K. (1979). The scoring of multiple-choice tests: A closer look. Educational Technology, 19, 5-15.

Anderson, R.I. (1982). Computer-based confidence testing: Alternatives to conventional, computer-based multiple-choice testing. Journal of Computer-Based Instruction, 9(1), 1-9.

Anderson, R.C., Kulhavy, R.W., & Andre, T. (1971). Feedback procedures in programmed instruction. Journal of Educational Psychology, 62, 148-156.

Carrol, J.B. (1976). Promoting language skills: The role of instruction. In D. Klahr Cognition and Instruction (pp. 323), Hillsdale, N.J.: Halsted Press.

Costermans, J., Lories, G., & Ansey, C. (1972). Confidence level and feeling of knowing in question answering: The weight of inferential processes. Journal of Experimental Psychology: Learning, Memory, and Cognition, 18(1), 142-150.

Crouse, R.W., Jr. (1980). The novelty of human self-assessment; Implications for learning and training. Unpublished master's thesis, New Mexico State University, Las Cruces, NM.

Eagly, A.H. (1978). Sex differences in influenceability. Psychological Bulletin, 85, 86-116.

Echternacht, G. (1972). The uses of confidence testing in objective tests. Review of Educational Research, 42, 217-236.

Fischoff, B., Slovic, P., & Lichenstein, S. (1977). Knowing with certainty: The appropriateness of extreme confidence. Journal of Experimental Psychology: Human Perception and Performance, 3, 552-564.

Fisher, P.M., & Mandl, H. (1985). Function and efficiency of contingent instrumental feedback. In G. d'Ydewalle (Ed.) Cognition, information processing and motivation. North Holland: Elsvier.

Henmon, V.A.C. (1911). The relation of the time of judgment to its accuracy. Psychological Review, 18, 186-201.

Hevner, K. (1932). A method of correcting for guessing in true-false tests and empirical evidence in support of it. Journal of Social Psychology, 3, 359-362.

Hollingsworth, H.L. (1913). Experimental studies in judgment. Archives of Psychology, 29, 1-119.

Hogarth, R.M., Gibbs, B.J., McKenzie, C.M., & Marquis, M.A. (1991). Learning from feedback: Exactingness and incentives. Journal of Experimental Psychology: Learning, Memory, and Cognition, 17(4), 734-752.

Hull, C.L. (1943). Principles of behavior. An introduction to behavior theory. New York: Appleton-Century-Crofts.

Hull, C.L. (1952). A behavior system: An introduction to behavior theory concerning the individual organism. New Haven, Conn.: Yale University Press.

Hunt, D.P. (1978). A human self-assessment process (A final report). Alexandria, VA.: U.S. Army Research Institute for the Behavioral and Social Sciences.

Hunt, D.P. (1982). Effects of human self-assessment responding on learning. Journal of Applied Psychology, 67, 75-82.

Hunt, D.P. (1984). Human self-assessment and its implications for human training and performance. Alexandria, VA.: U.S. Army Research Institute for the Behavioral and Social Sciences.

Hunt, D.P., & Sams, M.R. (1989). Human self-assessment process theory: An eight-factor model of human performance and learning; and everyman's causation. In G. Ljunggren & S. Dornic (Eds.), Psychophysics in action (pp. 42-53). New York: Springer-Verlag.

Hyde, J. S. (1981). How large are cognitive gender differences? A meta-analysis using w2 and d. American Psychologist, 36, 982-901.

Jacobs, S. S. (1971). Correlates of unwarranted confidence in responses to objective test items. Journal of Educational Measurement, 8, 15-19.

Johnston, V.S. (1979). Stimuli with biological significance. In Beglighter, H. (Ed.), Evoked brain potentials and behavior. New York, N.Y.: Plenum Press.

Kohlers, P.A., & Palef, S.R. (1976). Knowing not. Memory & Cognition, 4(5), 553-558.

Kulhavey, R.W., Yekovitch, F.R. & Dyer, J.W. (1976). Feedback and response confidence. Journal of Educational Psychology, 68, 522-528.

Meyer, S.R. (1960). Report on the initial test of a junior high school vocabulary program. In A.A. Lumsdaine & R. Glaser (Eds.), Teaching machines and programmed learning. Washington, D.C.: National Education Association.

Phye, G.D., & Bender, T. (1989). Feedback complexity and practice: Analysis in retention and transfer. Contemporary Education Psychology, 14, 97-110.

Proctor, R.W., Reeve, T.G., Weeks, D.J., Dornier, L., & Zandt, T.V. (1991). Acquisition, retention and transfer of response selection skill in choice reaction tasks. Journal of Experimental Psychology: Learning, Memory, and Cognition, 17(3), 497-506.

Rippey, R.M., & Voytovich, A.E. (1983). Linking knowledge, realism and diagnostic reasoning by computer-assisted confidence testing. Journal of Computer-Based Instruction, 9, 88-97.

Sams, M.R. (1986). Effects of testing methods and self-assessment responding on observers and performers. Unpublished master's thesis, New Mexico State University, Las Cruces.

Shaughnessy, J.J. (1979). Confidence-judgment accuracy as a predictor of test performance. Journal of Research in Personality, 13, 505-514.

Shuford, E.H. Jr., Albert, A., & Massengill, H.E. (1966). Admissible probability measurement procedures. Psychometrika, 31(2), 125-145.

Skinner, B.F. (1938). The behavior of organisms. New York: Appleton-Century.

Sutton, S., Braren, M., Zubin, J., & John, E.R. (1965). Evoked-potential correlates of stimulus uncertainty. Science, 150, 1187-1188.

Thorndike, E.L. (1898). Animal intelligence: An experimental study of the associative processes in animals. Psychological Review Monograph Supplements, 2(4), 109, [21-22, 54, 351].

Trow, W.C. (1923). The psychology of confidence. Archives of Psychology, 65, 1-47.

Woodson, W.E., & Tillman, B. (1992). Human factors design handbook; Information and guidelines for the design of systems, facilities, equipment and products for human use. 2d ed., NY: McGraw-Hill.

APPENDIX A

Pilot Study

Pilot Study

To evaluate and develop the experimental procedures a pilot study was conducted. Subjects were required to learn and apply the rules for identifying and extinguishing the four classes of fire to an 80% correct criterion. The experimental group overtly self-assessed the confidence they felt in the correctness of their response. This self-assessment was immediately followed by feedback on the correctness of the response. The control group was not required to assess the correctness of their response. Learning was measured by the number correct on a 30 question multiple choice task. After achieving the 80% criterion during the learning trial, the subject was required to return one week later at the same date and time to take an identical multiple choice task to evaluate retention.

Method

Subjects and Design

Thirty-seven college students served as subjects to partially satisfy requirements in an introductory psychology course. The data of an additional eight subjects was unusable as they (six) failed to return for the retention portion of the experiment or (two) were unable to learn the task to the 80% correct criterion. The experimental group learned with overt self-assessment

responding and tested for retention using overt self-assessment responding. The control group learned and was tested for retention without self-assessment responding (Appendix Table A-1).

Appendix Table A-1

Pilot Study Experimental Design

	HISPANIC AND NATIVE AMERICAN	WHITE NON-HISPANIC	TOTAL N
SA	13	6	19
NOSA	10	7	17

NOTE: SA = SELF-ASSESSMENT
NOSA = NO SELF-ASSESSMENT

Primary Learning Task

The primary task of all subjects was to learn the four classes of fire, the rules to determine which type of fire fell under which class, and the type

of fire extinguisher required to combat each class. A computer-generated tutorial providing the requisite information was displayed on the terminal screen for approximately 60 seconds for each class of fire. Computer-generated instructions for testing were presented only once during the first learning trial in a self-paced format with sample questions available for both self-assessors and non-self-assessors.

For the experimental group the self-assessment scoring rules were also explained and demonstrated with a sample question. Feedback on the correct response was provided for the experimental group immediately following their self-assessment of the correctness of the executed response and for the control group immediately following execution of the response.

Apparatus

The apparatus employed was a personal computer. All iterations of both self-assessment and non-self-assessment testing were performed on the same computer.

All relevant aspects of the experimental work station (i.e., dimensions of chair and learning station, visual angle, etc.) were in accordance with ergonomic principles determined by the Human Factors Design Handbook (Woodson & Tillman, 1992).

Procedure

Twenty subjects were randomly assigned to each of the two treatment groups (Appendix Table A-1). Each subject participated separately. The experimental conditions were standardized for day-of-week and time-of-day.

For all groups, subjects were told that the task was to learn the four classes of fire, and how to recognize and extinguish each class. They were instructed to study the information when presented to them because they would be tested on the information with a 30 question multiple choice test. They were informed that the standard for having learned the material was a score of 80% correct. See Appendix D for detailed instructions.

During both learning and retention trials, all subjects selected a response (A,B,C,or D) by pressing the corresponding pushbutton on the keyboard. The control group (NOSA) received immediate feedback on the correctness of their answer. The experimental group (SA) was first required to express the confidence felt in the accuracy of answer by selecting a self-assessment category ranging from 1-5 ("not sure at all" to "extremely sure"). Upon selection of this category, each group member received feedback on the correctness of their answer.

After successfully the completing the learning portion of the experiment, subjects were dismissed and instructed to return for 1 iteration of

an identical 30 question multiple choice task to evaluate retention the following week on the same day of the week at the same time. A sample of the multiple choice task is provided in Appendix E. The standard self-assessment question format that followed every test question for the experimental group is found in Appendix D in the test instructions.

Results

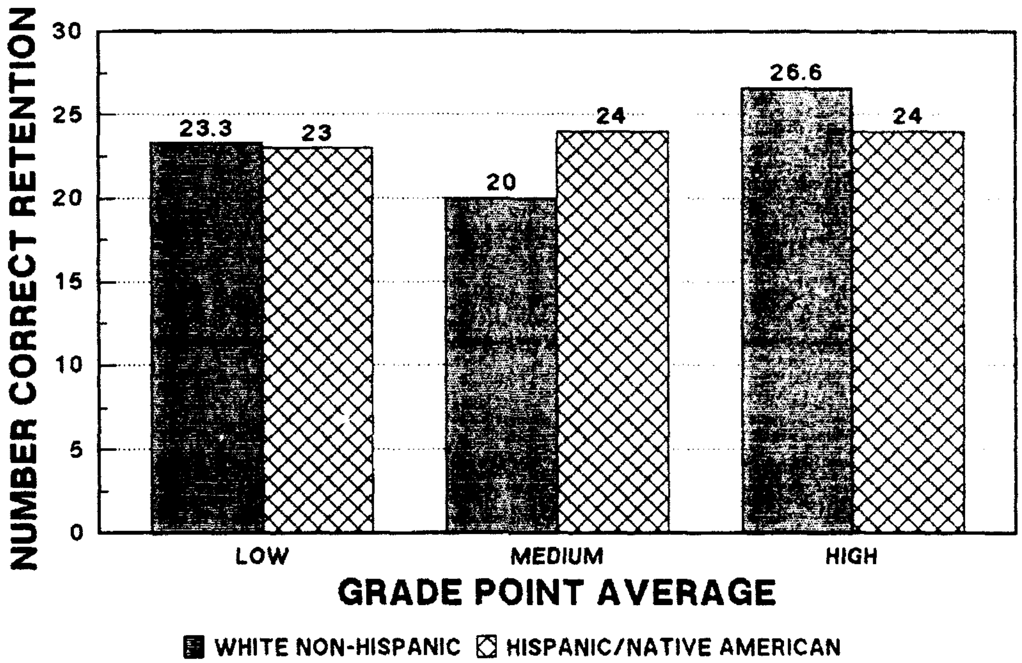
The number of correct responses was calculated for the final learning trial and the retention trial for each individual in both the control and experimental groups. It should be noted that the power of all tests performed was low due to the small sample size. There were 19 subjects that self-assessed and 17 non-self-assessors. The ethnic distribution was 6 White Non-Hispanics, 12 Hispanics, and 1 Native American that self-assessed and 7 White Non-Hispanics, 9 Hispanic and 1 Native American non-self-assessors.

An analysis of variance was conducted using the following four covariates: Self-assessment (SA) or No Self-assessment (NOSA), Gender, Ethnicity, and Grade Point Average (GPA). The dependent variable was the number correct on the retention trial.

Grade point average was separated into three categories (a) High = 3.5 to 4.0, (b) Med = 3.0 to 3.4, and (c) Low = 2.5 to 2.9. Due to the demographics of the Las Cruces, NM area and small sample size, ethnic groups were collapsed into two categories. Blacks and Others were not represented so the two groups were (a) Ethnic 1 = White Non-Hispanic; and (b) Ethnic 2 = Hispanic/Native American. As there were only 2 Native Americans represented in the study, they were included with the Hispanics for the purpose of analysis. While not culturally identical, combining these two groups provided a white/non-white contrast.

There was a significant interaction of ethnicity and grade point average ($p < .05$). Hispanics and Native Americans showed little effect due to GPA. White Non-Hispanics demonstrated a positive effect on the number of correct responses retained with a high GPA ($M = 26.6$) and a negative effect with a medium GPA ($M = 20$) (Appendix Figure A-1).

Protected Least Significant Difference analysis ($p = .05$) was performed on the means of the ethnicity by GPA interaction to attempt to isolate the significant effects.



Appendix Figure A-1. Mean number correct on retention trial for Hispanics and White Non-Hispanics by GPA level.

All possible comparisons were computed and only the following comparisons were significant: (a) White Non-Hispanics with a high GPA (3.5 to 4.0) scored significantly higher on retention tests than Hispanics/Native Americans at all three GPA levels; and (b) White

Non-Hispanics with a medium GPA (3.0 to 3.4) scored significantly lower than Hispanics/Native Americans at all three GPA levels.

White Non-Hispanics with a low GPA (2.5 to 2.9) did not score significantly higher or lower than Hispanics/Native Americans at all three GPA levels.

Analysis concerning the relation between (a) number correct on the final learning trial that were also retained correctly on the retention trial and (b) the self-assessment confidence level assigned to those correct responses during the final learning trial could not be performed due to the low number of observations per cell. To better illustrate the data, the number of correct responses is shown as a function of level of confidence indicated by the subject during the final learning trial (Appendix Figure A-2).

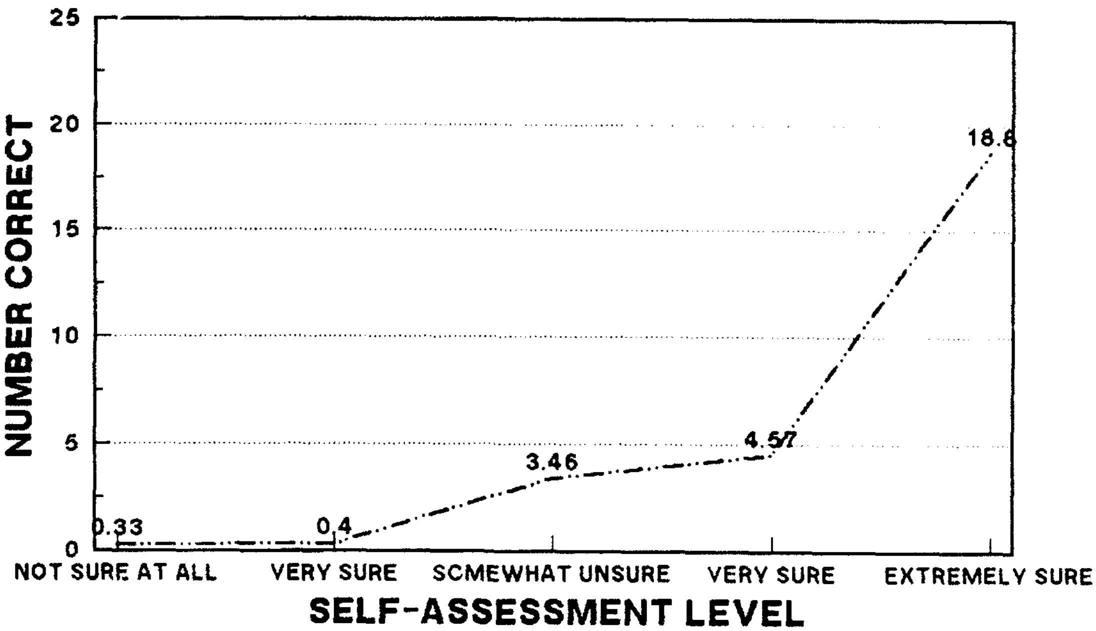
Observational data shown in Appendix Figure A-2 suggests that the more confident a subject is in the accuracy of his response, the greater the number of correct responses he will retain.

Discussion

Self-Assessment

The findings of this pilot study do not support the hypothesis that overt self-assessment responding while learning and testing increases the number correct on final learning and retention trials. However, the present

study suggests that the secondary hypothesis that responses about which an examinee is extremely sure and correct will be retained at a higher percentage than those responses about which he is less confident may be supported.



Appendix Figure A-2. Mean number of correct responses on retention test as a function of the confidence level indicated on final learning trial.

Further study is required to determine the extent of this relationship between sure-correct responses and retention of training.

Ethnicity and Grade Point Average

A significant interaction was found between grade point average (GPA) and ethnicity. Both White Non-Hispanics and Hispanics/Native Americans with the highest GPA level scored higher on retention trials. The significance was concentrated at the high end of the GPA scale. The difference between the two ethnic group's mean scores at the lowest GPA level is only 0.3.

The data indicate that GPA has a greater effect on White Non-Hispanic retention scores as they ranged from 26.6 at the highest GPA level to 20 at the medium level. Hispanic/Native American scores differed by only point from the high GPA level score of 24 to the low GPA level score of 23.

Hispanics/Native Americans at the low GPA level out scored White Non-Hispanics at the medium GPA level by three points and scored only 0.3 points less than White Non-Hispanics at the low GPA level. It seems that GPA is not related to the retention scores of Hispanics/Native Americans as there is only a one point difference in mean retention scores between low and high level GPA's for this group.

Further study is required to determine whether the higher retention scores of both White Non-Hispanics and Hispanics/Native Americans at highest GPA level are a result of the better test-taking strategies of subjects with a high GPA and to determine the effects of ethnicity on test-taking ability in conjunction with the use of self-assessment responding. Primary focus must be given to relationship between the confidence level assigned by the subject during the final learning trial and the accuracy of retention.

APPENDIX B

ERGONOMICS OF EXPERIMENTAL WORK STATION

Apparatus Design

Lateral location for hand control. Recommended: Hand controls in frequent or continuous use should be located in front of the operator in the area between the shoulders, an area of about 16-18 inches wide. Actual : Keyboard was located in front of the operator, between the shoulders. Length of keyboard was 17 inches.

Controls. Recommended: Push button separation should be one inch for one finger used sequentially. Actual: Pushbuttons were 16 cm wide, 24 cm apart and square shaped.

Viewing distance. Recommended: viewing distance to displays with reaching controls should never be less than 13 inches or more than 29.5 inches from the eyes for vertical panels. Actual: Distance from the bottom of the keyboard panel was 21 inches and 23 inches from the top of the panel. Distance from the monitor was 19 inches.

Viewing angle. Recommended: When the head and eyes can be rotated, the optimum lateral angle is 15 degrees to the left and 15 degrees to the right. The optimum vertical angle is 0 degrees up and 30 degrees down. Actual: Lateral viewing is 15 degrees to the right and to the left. Vertical viewing angle is 0 degrees and 35 degrees down.

Seat Design

Seat pan. Recommended: seat should be 15 to 16 inches above the floor. The best general purpose seat length is about 17 inches. A reasonable

dimension for seat width is 18 inches. Actual: Seat pan was 18 inches above the floor, 18 inches wide, and 18.5 inches long.

Chair backrest. Recommended: Backrest should extend at least 18 to 20 inches above the seat pan to provide back support up to the shoulder area. At least a 20 inch wide backrest will provide full support across the shoulders. Actual: Backrest extended 16 inches above the seat pan with a 17 inch wide backrest.

Armrests. Recommended: Height should be 8 to 10 inches above the seat pan. Actual: Height was 9 inches above the seat pan.

Table

Table height. Recommended: Table height should be in the same plane as the armrest of the chair. This supports the arm without forcing the operator to raise or depress the shoulder. Actual: Table was 26.5 inches high and the armrest was less than inch above the table.

APPENDIX C

Representative Experimentation Schedule

Appendix Table C-1
Self-Assessment Scoring

	POINTS AWARDED FOR CORRECT RESPONSE	POINTS AWARDED FOR WRONG RESPONSE
SA LEVEL		
NOT SURE AT ALL	+10	+5
VERY UNSURE	+27	-4
SOMEWHAT UNSURE	+37	-16
VERY SURE	+45	-32
EXTREMELY SURE	+50	-60

NOTE:

Although not used as the experimental dependent measure, these values can be evaluated for a percent correct score during training.

APPENDIX D

Instructions to Subjects

No Self-Assessment

A COMPUTER ASSISTED TEST

This test requires you to select the ANSWER from among the alternatives tested. One benefit of this test is that you receive immediate feedback as to whether your answer is correct. Your test score will depend on

- the CORRECTNESS of your answer.

During the test, a question will be presented. You should try to answer it as well as you can.

For example, you might be asked to select the capitol of Ohio from among three cities (A, B, or C).

A. Dayton

B. Colombus

C. Covington

You might want answer B (it's correct). You would do this by pressing the "B" key on the computer keyboard - then pressing the "ENTER" button.

If you press a wrong letter you can erase it (before you press Enter) by pressing the "backward arrow" (<==) key on the top row.

You GAIN points for giving a CORRECT ANSWER and

- you LOSE points by giving a WRONG ANSWER.

Your score on each question is based on the following scale:

	POINTS
CORRECT ANSWER	+ 50

WRONG ANSWER - 60

since your answer of Colombus is CORRECT-- you will gain 50 points.

However, if your answer is WRONG

-- you will lose 60 points (instead of only 50).

So...

the more correct answers you give, the higher your score on the test.

The particular point assignments have been selected so that

YOU WILL OBTAIN THE HIGHEST TEST SCORE

by

SELECTING YOUR ANSWERS AS CAREFULLY AS YOU CAN

Would you like a practice question?

(YES or NO, then press ENTER)

Self-Assessment

On this test you

first, select an ANSWER

then, indicate HOW SURE YOU ARE that your answer is correct.

Your test score depends on

1. the CORRECTNESS of your answer, and
2. you can obtain bonus points for the ACCURACY of your confidence assessment.

For example, you might be asked to select the capital of Ohio from among four cities (A, B, C, or D).

A. Dayton
B. Columbus
C. Covington
D. Cincinnati

You might want to answer B (it's correct). You would do this by pressing the "B" key on the computer keyboard - then pressing the Enter key.

After you give your answer, the following scale will be presented:

HOW SURE ARE YOU THAT YOUR ANSWER IS CORRECT?

1. Not Sure At All	2. Very Unsure	3. Somewhat Unsure	4. Very Sure	5. Extremely Sure

---

If you are almost certain that Colombus is the correct answer, you would press the number "5" key, indicating that you are "EXTREMELY SURE" that it is correct -- then press the Enter key.

If you press the wrong letter or number, you can erase it, before you press the Enter key by pressing the "Backspace" key.

You get POINTS for giving a CORRECT ANSWER

You get BONUS POINTS for making an ACCURATE CONFIDENCE ASSESSMENT!

So .... the more accurate your confidence assessments,
.... the higher your score on the test.

The particular points for scoring have been selected so that

YOU WILL OBTAIN THE HIGHEST SCORE

by

ACCURATELY AND TRUTHFULLY INDICATING "HOW SURE" YOU ARE.

Would you like a practice question?

(YES or NO, then press Enter)

APPENDIX E
Test Questions

APPENDIX E

TEST QUESTIONS

1.) You are out burning leaves and your small, controlled burn becomes a raging inferno. What type of fire extinguisher do you use?

   a. Pump Tank
   b. Loaded stream
   c. Multi-purpose Dry Chemical
   d. Special Dry Chemical

2.) What type of fire extinguisher combats combustible metal fires?

   a. Loaded Stream
   b. Pump Tank
   c. Special Dry Chemical
   d. Multi-purpose Dry Chemical

3.) While working the graveyard shift at the Chemicals R' Us plant, you notice a fire in the sodium storage bin. What type of fire extinguisher do you use?

   a. Special Dry Chemical
   b. Multi-purpose Dry Chemical
   c. Loaded Stream
   d. Pump Tank

4.) What type of fire extinguisher would you use to combat a fire that requires blanketing or smothering?

   a. Pump Tank
   b. Special Dry Chemical
   c. Multi-purpose Dry Chemical
   d. Loaded Stream

5.) You are trying to record the final episode of Dallas when you notice a fire in the maze of wiring behind your T.V. What type of fire extinguisher do you use?

   a. Loaded Stream
   b. Multi-purpose Dry Chemical
   c. Pump Tank
   d. Special Dry Chemical

6.) What type of fire extinguisher would you use to combat a fire that requires special extinguishing agents or techniques?

a. Special Dry Chemical
b. Loaded Stream
c. Multi-purpose Dry Chemical
d. Pump Tank

7.) Your psychology term paper is due. While you're copying one final article, you notice flames coming out of the paper recycling bin. What type of fire extinguisher would you use?

a. Pump Tank
b. Multi-purpose Dry Chemical
c. Loaded Stream
d. Special Dry Chemical

8.) What type of fire extinguisher would you use to combat a fire that deals with electrical equipment?

a. Multi-purpose Dry Chemical
b. Loaded Stream
c. Special Dry Chemical
d. Pump Tank

9.) After frying yourself a midnight snack of french fries, you notice that there is a fire in the pan you were frying in. What type of fire extinguisher would you use?

a. Loaded Stream
b. Pump Tank
c. Special Dry Chemical
d. Multi-purpose Dry Chemical

10.) What type of fire extinguisher would you use to combat a fire that deals with ordinary combustible materials?

a. Pump Tank
b. Special Dry Chemical
c. Multi-purpose Dry Chemical
d. Loaded Stream

11.) While washing beakers in the chemistry lab, you notice a fire in the titanium container. What type of fire extinguisher would you use?

a. Multi-purpose Dry Chemical
b. Special Dry Chemical
c. Loaded Stream
d. Pump Tank

12.) What type of fire extinguisher would you sue to combat a fire that requires cooling, drenching?

a. Loaded Stream
b. Multi-purpose Dry Chemical
c. Pump Tank
d. Special Dry Chemical

13.) While playing the bass guitar in your band, you hear snapping sounds coming from your amp and you look down and sees sparks and smoke. What type of fire extinguisher would you use?

a. Pump Tank
b. Loaded Stream
c. Multi-purpose Dry Chemical
d. Special Dry Chemical

14.) What type of fire extinguisher would you use to combat a fire in ordinary packing materials?

a. Special Dry Chemical
b. Pump Tank
c. Loaded Stream
d. Multi-purpose Dry Chemical

15.) Hard at work at the Processed Foods Limited factory, you notice a fire in the potassium headed for the processing room. What type of fire extinguisher would you use?

a. Loaded Stream
b. Special Dry Chemical
c. Pump Tank
d. Multi-purpose Dry Chemical

16.) What type of fire extinguisher would you use to combat a fire in flammable liquids, grease?

a. Pump Tank
b. Special Dry Chemical
c. Loaded Stream
d. Multi-purpose Dry Chemical

17.) While painting the trim on Grandma's house, you look over and notice that a pile of rags which have been soaked in paint thinner, as well as a can of paint thinner itself have ignited. What type of fire extinguisher do you use?

a. Multi-purpose Dry Chemical
b. Loaded Stream
c. Special Dry Chemical
d. Pump Tank

18.) What type of fire extinguisher would you use to combat a fire that requires a non-conducting agent?

a. Loaded Stream
b. Multi-purpose Dry Chemical
c. Pump Tank
d. Special Dry Chemical

19.) While walking across campus, you notice that someone has tossed a lit match into a trashcan filled mostly with papers and wrappers. What type of fire extinguisher would you use?

a. Multi-purpose Dry Chemical
b. Pump Tank
c. Loaded Stream
d. Special Dry Chemical

20.) What type of fire extinguisher would you use to combat a fire in a control panel?

a. Loaded Stream
b. Special Dry Chemical
c. Pump Tank
d. Multi-purpose Dry Chemical

21.) As you attempt to ignite your gas grill to barbecue some steaks, your propane container catches on fire. What type of fire extinguisher do you use?

a. Pump Tank
b. Multi-purpose Dry Chemical
c. Loaded Stream
d. Special Dry Chemical

22.) While driving across the Mojave Desert, you notice your oil light flashing. You attempt to add a quart of oil and it ignites. What type of fire extinguisher do you use?

a. Special Dry Chemical
b. Loaded Stream
c. Multi-purpose Dry Chemical
d. Pump Tank

23.) Upon graduation from NMSU, you have been hired on in the bulk chemical department of a battery factory. Your first day on the job you encounter a fire in the magnesium assembly area. What type of fire extinguisher do you use?

a. Multi-purpose Dry Chemical
b. Special Dry Chemical
c. Pump Tank
d. Loaded Stream

24.) You are parked on Transmountain Drive and neglect to put on your emergency brake. As the tow truck winches your car out of the ravine, the winch catches on fire. What type of fire extinguisher do you use?

a. Special Dry Chemical
b. Loaded Stream
c. Multi-purpose Dry Chemical
d. Pump Tank

25.) After completing your 800th article of piece work for the Levi factory, you notice a fire in the reject pile. What type of fire extinguisher do you use?

a. Loaded Stream
b. Special Dry Chemical
c. Pump Tank
d. Multi-purpose Dry Chemical

26.) While plugging in your word processor, you notice sparks and flames coming from the socket. What type of fire extinguisher do you use?

a. Multi-purpose Dry Chemical
b. Loaded Stream
c. Special Dry Chemical
d. Pump Tank

27.) While driving to Arizona, you notice smoke coming from under your hood. After pulling over, you open the hood and flames leap out. What type of fire extinguisher do you use?

a. Pump Tank
b. Loaded Stream
c. Multi-purpose Dry Chemical
d. Special Dry Chemical

28.) Someone flicks a lit cigarette into the spare can of gasoline that you keep on hand for emergencies. What type of fire extinguisher do you use?

a. Special Dry Chemical
b. Multi-purpose Dry Chemical
c. Loaded Stream
d. Pump Tank

29.) While studying nonstop for 5 days during finals, you fall asleep on your bed with a lit cigarette in your hand. What type of fire extinguisher do you use?

a. Pump Tank
b. Loaded Stream
c. Multi-purpose Dry Chemical
d. Special Dry Chemical

30.) While checking your homemade still, you notice a fire in the grain alcohol container. What type of fire extinguisher do you use?

a. Special Dry Chemical
b. Multi-purpose Dry Chemical
c. Loaded Stream
d. Pump Tank

APPENDIX F

Self-Assessment Scoring

Appendix Table F-1

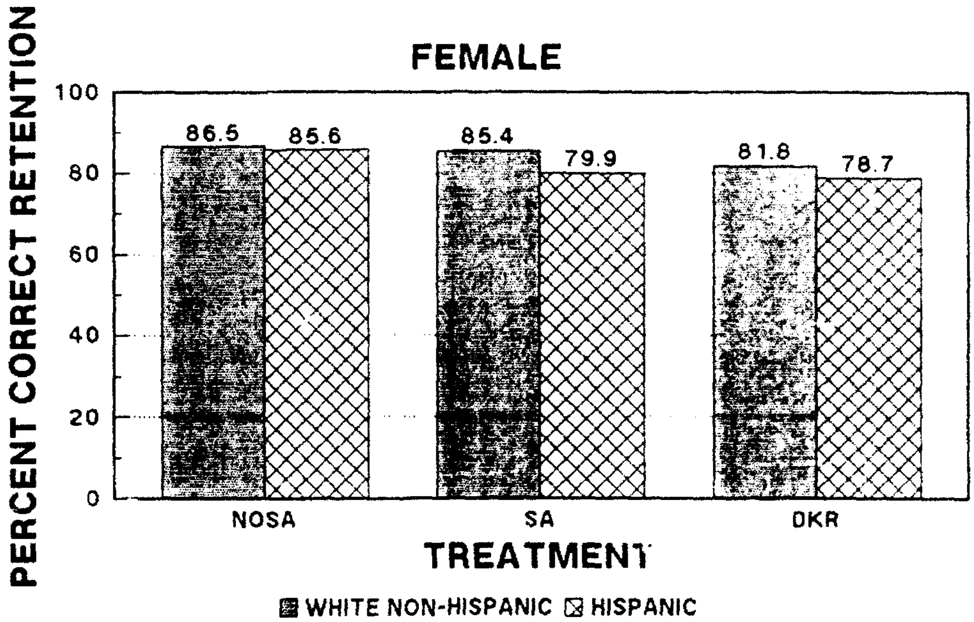
Self-Assessment Scoring

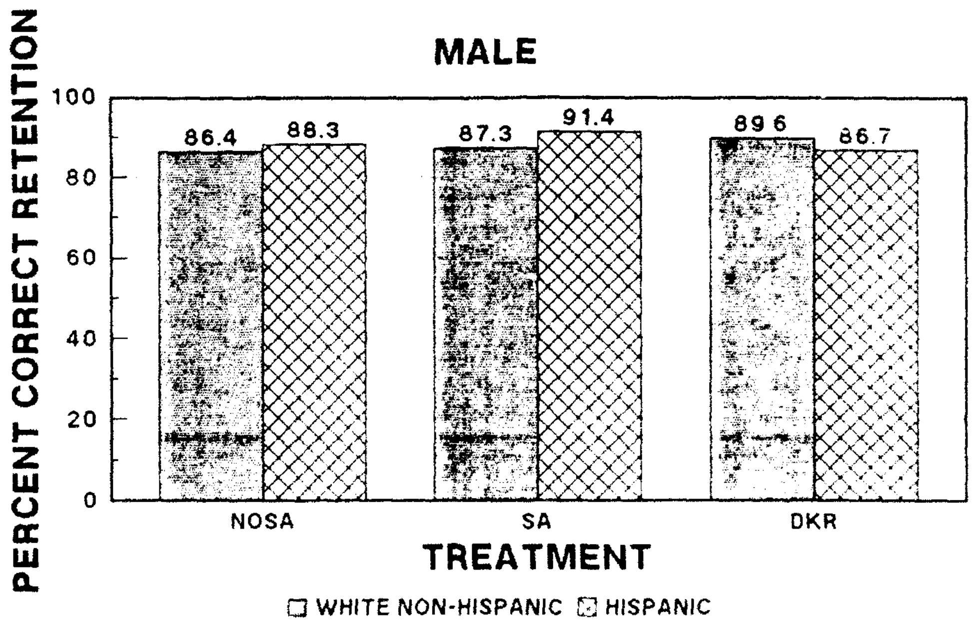
Self-Assessment Scale	**Points Awarded**	
	Correct Response	**Wrong Response**
1. Not Sure At All	+ 10	+ 5
2. Very Unsure	+ 27	- 4
3. Somewhat Unsure	+ 37	- 16
4. Very Sure	+ 45	- 32
5. Extremely Sure	+ 50	- 60

NOTE: Although not used as the experimental dependent measure, these values can be evaluated for a percent correct self-assessment score during training.

APPENDIX G

Supplementary Figures and Tables





Appendix Figure G-1. Mean percent correct retention trial by ethnicity, treatment, and gender.

Appendix Table G-1

Mean Percent Correct and Standard Deviation Retention by Treatment, Ethnicity, and Gender

	SA	NOSA	DKR	HISPANIC	WNH	MALE	FEMALE
MEAN	86	86.7	84.2	85.1	86.1	88.2	82.9
STD DEV	9.26	9.18	9.7	10.04	8.69	6.83	10.77

NOTE: SA = SELF-ASSESSMENT, NOSA = NO SELF- ASSESSMENT
DKR = DELAY KNOWLEDGE OF RESULTS, WNH = WHITE NON-HISPANIC

Appendix Table G-2
Analysis of Covariance by Treatment, Ethnicity, and Gender, Considering GPA as a Covariate

	DF	F VALUE	PR < F
GPA	2	1.25	0.2909
GROUP	2	0.85	0.4311
GENDER	1	11.51	0.0010
ETHNIC	1	0.08	0.7792
GROUP*GENDER	2	1.55	0.2161
GROUP*ETHNIC	2	0.60	0.5499
GENDER*ETHNIC	1	1.59	0.2104
GROUP*GENDER*ETHNIC	2	0.43	0.6538
TOTAL	13		

Appendix Table G-3
Number of Learning Trials to Reach 80% Correct by Treatment

	1 TRIAL	2 TRIALS	3 TRIALS	4 TRIALS	5 - 10 TRIALS	TOTAL N
SA	29	3	5	1	2	40
DKR	20	8	4	1	7	40
NOSA	14	9	6	5	6	40

NOTE: SA = SELF-ASSESSMENT NOSA = NO SELF-ASSESSMENT
DKR = DELAY KNOWLEDGE OF RESULTS

Appendix Table G-4

Cumulative Number of Relearning Trials to Reach 100% Correct for SA and NOSA Groups

CUMULATIVE TRIALS	SELF-ASSESSMENT	NO SELF-ASSESSMENT	DIFFERENCE
2	1	1	0
3	8	4	4
4	14	9	5
5	24	14	10
6	28	17	11
7	32	23	9
8	34	23	11
9	37	31	6
10	39	35	4
11	39	38	1
12	39	38	1
13	39	39	0
16	39	39	0
18	40	40	0
19	40	40	0

NOTE: SA = SELF-ASSESSMENT NOSA = NO SELF-ASSESSMENT

Appendix Table G-5

Cumulative Number of Relearning Trials to Reach 100% Correct for DKR and NOSA Croups

	DELAY KNOWLEDGE OF RESULTS	NO SELF-ASSESSMENT	DIFFERENCE
CUMULATIVE TRIALS			
2	1	1	0
3	8	4	4
4	13	9	4
5	19	14	5
6	28	17	11
7	29	23	6
8	32	23	9
9	34	31	3
10	35	35	0
11	36	38	2
12	38	38	0
13	38	39	1
16	39	39	0
18	39	40	1
19	40	40	0

NOTE: DKR = DELAY KNOWLEDGE-OF-RESULTS NOSA = NO SELF-ASSESSMENT

Appendix Table G-6

Number of Final Learning Trial Correct Responses Retained Correctly

	SELF-ASSESSMENT LEVEL 1 AND 2	SELF-ASSESSMENT LEVEL 3	SELF-ASSESSMENT LEVEL 4	SELF-ASSESSMENT LEVEL 5	TOTAL NUMBER RESPONSES
NUMBER RETAINED CORRECTLY	7	40	85	835	967
NUMBER CORRECT LEARNING TRIAL	12	53	97	915	1,077